Econometria
Modelo de Regresión Lineal Simple

Pasquini, Ricardo

IAE Business School
Universidad Austral

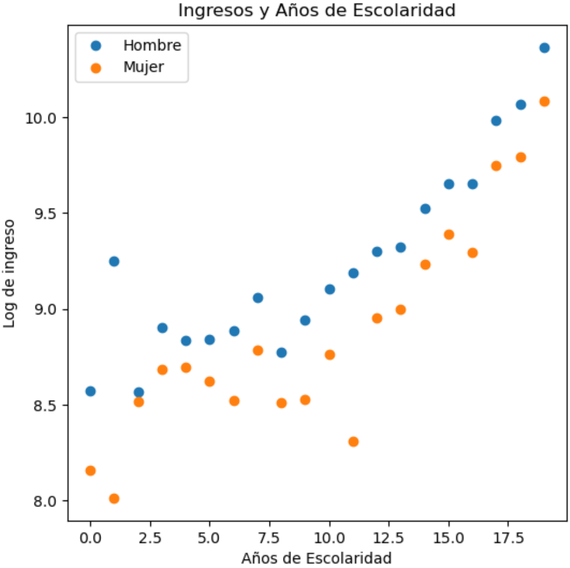
20 de agosto de 2025

Objetivos de Aprendizaje

- Definir un modelo de regresión con una variable explicativa.
- Comprender los componentes del modelo de regresión.
- Aprender a estimar el modelo de regresión.
- Interpretar los coeficientes de regresión.
- Evaluar la bondad de ajuste del modelo.

Motivación: Modelo CEF

Ejemplo: Ingresos y Años de Escolaridad



Modelo de regresión simple

$$\underbrace{Y_i}_{\substack{\text{Variable}\\\text{a}\\\text{explicar}\\\text{o}\\\text{Dependiente}}} = \underbrace{\beta_0}_{\substack{\text{Constante}\\\text{o inter-}\\\text{cepto}}} + \beta_1 \cdot \underbrace{X_i}_{\substack{\text{Variable}\\\text{explicativa}\\\text{1}\\\text{(independiente)}}} + \underbrace{\varepsilon_i}_{\text{Error}}$$

- β_0, β_1 son los coeficientes a ser estimados.
- Una vez estimados los denotamos con $\hat{\beta}_0, \hat{\beta}_1$.
- $\hat{\beta}_0$ es el intercepto y $\hat{\beta}_1$ es la pendiente.

Estimador de Mínimos Cuadrados Ordinarios (OLS)

- El estimador OLS es utilizado para estimar los coeficientes de un modelo de regresión lineal.
- Minimiza la suma de los cuadrados de los residuos (RSS), que es la diferencia entre los valores observados de Y y los valores predichos por el modelo.

Estimador de Mínimos Cuadrados Ordinarios (OLS)

Paso 1: Definir la función de pérdida (RSS):

$$RSS = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

Paso 2: Minimizar la función de pérdida en términos de β_0 y β_1.

$$\min_{\beta_0, \beta_1} RSS = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2$$

Paso 3: Resolver las ecuaciones resultantes para β_0 y β_1.

Estimador de Mínimos Cuadrados Ordinarios (OLS)

Derivadas Parciales:

Derivamos la función de pérdida (RSS) respecto a β_0 y β_1 y las igualamos a cero para minimizar RSS:

$$\frac{\partial RSS}{\partial \beta_0} = -2 \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i) = 0$$

$$\frac{\partial RSS}{\partial \beta_1} = -2 \sum_{i=1}^{n} X_i (Y_i - \beta_0 - \beta_1 X_i) = 0$$

Estas ecuaciones conducen a las soluciones para $\hat{\beta}_0$ y $\hat{\beta}_1$.

Estimador de Mínimos Cuadrados Ordinarios (OLS)

Las ecuaciones resultantes para $\hat{\beta}_0$ y $\hat{\beta}_1$ son:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}$$

Donde $\bar{X}$ y $\bar{Y}$ son las medias muestrales de X y Y respectivamente.

Estos son los estimadores de mínimos cuadrados ordinarios (OLS) para el modelo de regresión lineal simple.

Estimadores de Mínimos Cuadrados Ordinarios (OLS)

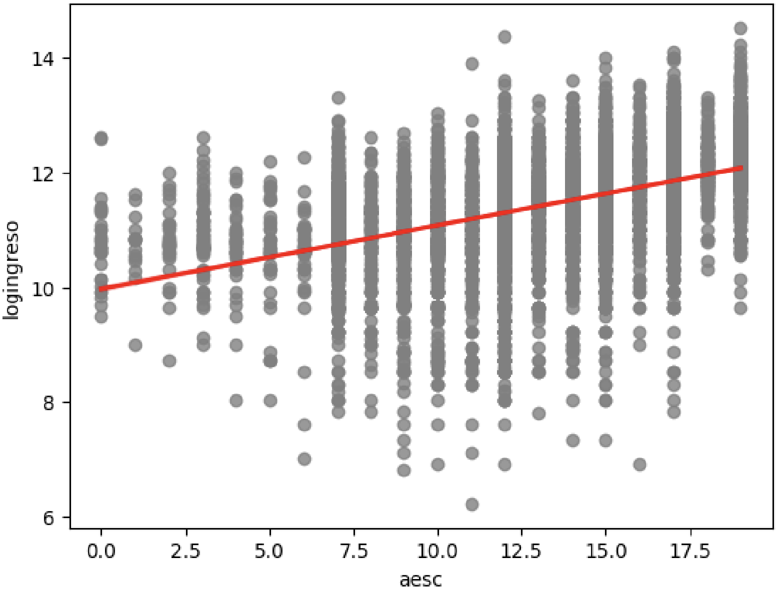
Ejemplo: Ingresos y Años de Escolaridad

```
modelo=smf.ols(formula='logingreso ~ aesc',data=df2)
resultados=modelo.fit()
print(resultados.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             logingreso   R-squared:                       0.207
Model:                            OLS   Adj. R-squared:                  0.207
Method:                 Least Squares   F-statistic:                     2366.
Date:                Wed, 20 Aug 2025   Prob (F-statistic):               0.00
Time:                        17:59:52   Log-Likelihood:                -10880.
No. Observations:                9062   AIC:                         2.176e+04
Df Residuals:                    9060   BIC:                         2.178e+04
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      9.9699      0.031    320.835      0.000       9.909      10.031
aesc           0.1104      0.002     48.639      0.000       0.106       0.115
==============================================================================
Omnibus:                     1489.077   Durbin-Watson:                   1.822
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             3561.197
Skew:                          -0.933   Prob(JB):                         0.00
Kurtosis:                       5.440   Cond. No.                         50.6
```

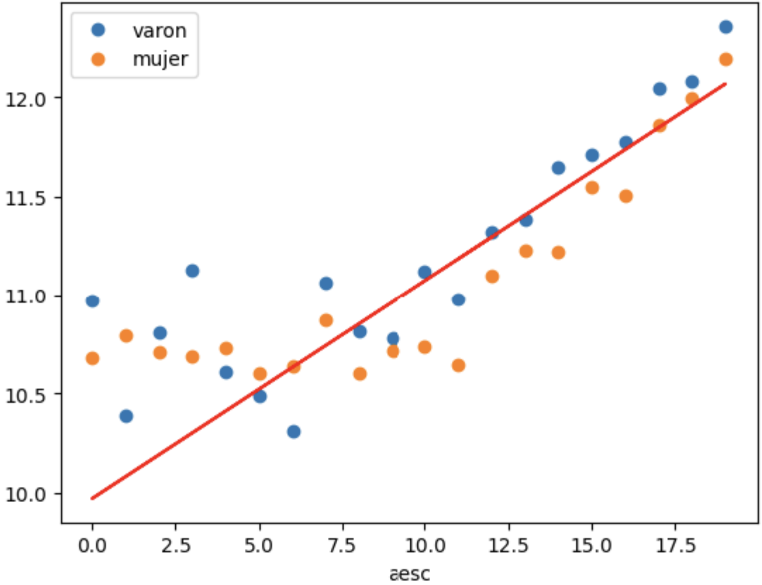
Estimadores de Mínimos Cuadrados Ordinarios (OLS)

Ejemplo: Ingresos y Años de Escolaridad



Estimadores de Mínimos Cuadrados Ordinarios (OLS)

Ejemplo: Regresión y CEF



Bondad de Ajuste

- Medidas de bondad de ajuste: R^2 y MSE.

Bondad de Ajuste - R^2

- El coeficiente R^2 se define como la proporcion de la varianza de la variable dependiente Y que es explicable por la variable X.
- Paso 1: Calcular los valores predichos de Y utilizando la ecuación de regresión:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

  Donde $\hat{Y}_i$ es el valor predicho de Y para la i-ésima observación, y $\hat{\beta}_0$ y $\hat{\beta}_1$ son los coeficientes estimados obtenidos del análisis de regresión.
- Paso 2: Calcular la suma total de cuadrados (TSS), que mide la variabilidad total en la variable dependiente:

$$TSS = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

Bondad de Ajuste - R^2

- Paso 3: Calcular la suma de cuadrados residual (RSS), que mide la variabilidad que no es explicada por el modelo de regresión:

$$RSS = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

- Paso 4: Calcular la suma de cuadrados explicada (ESS), que mide la variabilidad explicada por el modelo de regresión:

$$ESS = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2$$

- Paso 5: Calcular R^2 utilizando la fórmula:

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

Bondad de Ajuste - R^2

```
modelo=smf.ols(formula='ingtot_2 ~ aesc',data=df2)
resultados=modelo.fit()
print(resultados.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:               ingtot_2   R-squared:                       0.263
Model:                            OLS   Adj. R-squared:                  0.263
Method:                 Least Squares   F-statistic:                     4437.
Date:                Wed, 20 Aug 2025   Prob (F-statistic):               0.00
Time:                        15:27:36   Log-Likelihood:            -1.6070e+05
No. Observations:               12432   AIC:                         3.214e+05
Df Residuals:                   12430   BIC:                         3.214e+05
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept  -3.578e+04   2150.736    -16.638      0.000      -4e+04   -3.16e+04
aesc        1.149e+04    172.493     66.608      0.000    1.12e+04    1.18e+04
==============================================================================
Omnibus:                    10130.562   Durbin-Watson:                   1.837
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           559375.099
Skew:                           3.527   Prob(JB):                         0.00
Kurtosis:                      35.095   Cond. No.                         30.2
==============================================================================
```

Error Cuadrático Medio (MSE)

- El MSE es una medida de la calidad de un modelo de regresión.
- Indica el promedio de los cuadrados de los errores, es decir, la diferencia entre los valores predichos por el modelo y los valores reales de la variable dependiente.

Error Cuadrático Medio (MSE)

Fórmula

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

- Donde Y_i es el valor real de la variable dependiente para la i-ésima observación.
- $\hat{Y}_i$ es el valor predicho por el modelo de regresión para la i-ésima observación.
- n es el número total de observaciones.

Error Cuadrático Medio (MSE)

- **Paso 1:** Calcular los residuos para cada observación:

$$\hat{\varepsilon}_i = Y_i - \hat{Y}_i$$

- **Paso 2:** Elevar al cuadrado cada error:

$$\hat{\varepsilon}_i^2 = (Y_i - \hat{Y}_i)^2$$

- **Paso 3:** Calcular el promedio de los residuos al cuadrado:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} \hat{\varepsilon}_i^2$$

Referencias

- Wooldridge ch. 2